

# Exhibit 1

LOCUS AY057051 2245 bp mRNA linear PRI 11-AUG-2003  
 DEFINITION Homo sapiens lysosomal associated transmembrane protein 4 beta, variant 1 (LAPTM4B) mRNA, complete cds.  
 ACCESSION AY057051  
 VERSION AY057051.2 GI:30146628  
 KEYWORDS .  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;  
 Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;  
 Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 2245)  
 AUTHORS Shao, G. Z., Zhou, R. L., Zhang, Q. Y., Zhang, Y., Liu, J. J., Rui, J. A., Wei, X. and Ye, D. X.  
 TITLE Molecular cloning and characterization of LAPTM4B, a novel gene upregulated in hepatocellular carcinoma  
 JOURNAL Oncogene 22 (32), 5060-5069 (2003)  
 PUBMED 12902989  
 REFERENCE 2 (bases 1 to 2245)  
 AUTHORS Shao, G. and Zhou, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-SEP-2001) Cell Biology, Peking University Health Science Center, 38# XueYuan Road, Beijing 100083, China  
 REFERENCE 3 (bases 1 to 2245)  
 AUTHORS Shao, G. and Zhou, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (23-MAY-2002) Cell Biology, Peking University Health Science Center, 38# XueYuan Road, Beijing 100083, China  
 REMARK Sequence update by submitter  
 REFERENCE 4 (bases 1 to 2245)  
 AUTHORS Shao, G. and Zhou, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-MAR-2003) Cell Biology, Peking University Health Science Center, 38# XueYuan Road, Beijing 100083, China  
 REMARK Sequence update by submitter  
 COMMENT On Apr 28, 2003 this sequence version replaced gi:16508140.

## A Practical Guide to the Analysis of Genes and Proteins

EDITED BY  
ANDREAS D. BAXEVANIS  
B. F. FRANCIS OUELLETTE

WWW.

The third item on the locus line indicates the molecule type. The "mol type" usually is DNA or RNA, and it can also indicate the strandedness (single or double, as ss or ds, respectively); however, these attributes are rarely used these days (another historical leftover). The acceptable mol types are DNA, RNA, tRNA, rRNA, mRNA, and uRNA and are intended to represent the original biological molecule. For example, a cDNA that is sequenced really represents an mRNA, and mRNA is the indicated mol type for such a sequence. If the tRNA or rRNA has been sequenced directly or via some cDNA intermediate, then tRNA or rRNA is shown as the mol type. If the ribosomal RNA gene sequence was obtained via the PCR from genomic DNA, then DNA is the mol type, even if the sequence encodes a structural RNA.

The fourth item on the locus line is the GenBank division code: three letters, which have either taxonomic inferences or other classification purposes. Again, these codes exist for historical reasons, recalling the time when the various GenBank divisions were used to break up the database files into what was then a more manageable size. The GenBank divisions are slightly different from those of EMBL or DDBJ, as described elsewhere (Ouellette and Boguski, 1997). NCBI has not introduced additional organism-based divisions in quite a few years, but new, function-based divisions have been very useful because they represent functional and definable sequence types (Ouellette and Boguski, 1997). The Expressed Sequence Tags (EST) division was introduced in 1993 (Boguski et al., 1993) and was soon followed by a division for Sequence Tagged Sites (STS). These, along with the Genome Survey Sequences (GSS) and unfinished, High Throughput Genome sequences (HTG), represent functional categories that need to be dealt with by the users and the database staff in very different ways. For example, a user can query these data sets specifically (e.g., via a BLASTN search against the EST or HTG division). Knowing that the hit is derived from a specific technique-oriented database allows one to interpret the data accordingly. At this time, GenBank, EMBL, and DDBJ interpret the various functional divisions in the same way, and all data sets are represented in the same division from one database to the next. The CON division is a new division for constructed (or "contigged") records. This division contains segmented sets as well as all large assemblies, which may exceed (sometimes quite substantially) the 350,000-bp limit presently imposed on single records. Such records may take the form shown in Appendix 3.3. The record from the CON division shown in Appendix 3.3 gives the complete genomic sequence of *Mycoplasma pneumoniae*, which is more than 800,000 base pairs in length. This CON record does not include sequences or annotations; rather, it includes instructions on how to assemble pieces present in other divisions into larger or assembled pieces. Records within the CON division have accession and version numbers and are exchanged, like all other records within the collaboration.

The date on the locus line is the date the record was last made public. If the record has not been updated since being made public, the date would be the date that it was first made public. If any of the features or annotations were updated and the record was rereleased, then the date corresponds to the last date the entry was released. Another date contained in the record is the date the record was submitted (see below) to the database. It should be noted that none of these dates is legally binding on the promulgating organization. The databases make no claim that the dates are error-free; they are included as guides to users and should not be submitted in any arbitration dispute. To the authors' knowledge, they have never been used in establishing priority and publication dates for patent application.